METODOS DE PREVISIÓN

Obtener previsiones es asignar probabilidades a acontecimientos futuros .

* Métodos no causales : Se basan en el supuesto de que dada una serie histórica , es posible prever los valores futuros de la serie sin necesidad de examinar los factores que la han generado .

* Métodos causales : Relacionan los cambios de los valores de la variable a prever con los cambios de otras variables que se denominan variables explicativas . Dentro de éste método se utilizará la técnica de " regresión " .

MODELO LINEAL DE REGRESIÓN SIMPLE

Este modelo estudia la relación existente entre una var. explicada (y) y una var. explicativa (x) , ambas var. cuantitativas .
El modelo relaciona ambas var. mediante " la recta de regresión poblacional "

$$Y_i = A + BX_i + E_i$$

donde A y B son parámetros desconocidos del modelo , que deberán ser estimados a través de los datos provenientes de la muestra .

Forma de estimar estos parámetros : $\widehat{Y}_i = a + b\,x_i$ (recta de regresión muestral)

* Mediante el diagrama de dispersión : La representación gráfica de la nube de puntos dará una idea aprox. del tipo de relación . (lineal , parabólica ,etc)

* Mediante el metodo de los minimos cuadrados .
  La suma de los cuadrados de los residuos deberá ser minima .

$$S = \Sigma\,(y_i - a - bx_i)^2$$

y_i = valor observado $\widehat{y}_i = a + bx_i$ valor dado por la recta .

$$\partial S / \partial a = 0 \qquad \partial S / \partial b = 0$$

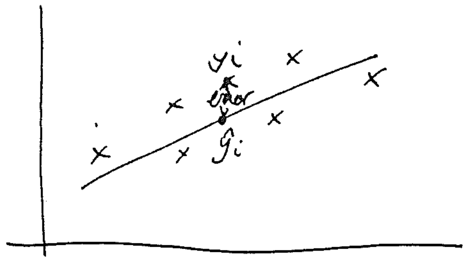
$$b = \frac{\Sigma X_i Y_i - n\overline{X}\,\overline{Y}}{\Sigma X_i^2 - n\overline{X}^2} \qquad a = \overline{Y} - b\overline{X}$$

$E(b) = B$ a estima puntualmente A
$E(a) = A$ b estima puntualmente B

* Propiedades de los estimadores minimocuadráticos.

$$a \rightarrow N\left(A ; S^2_{yx} \left[1/n + \frac{\overline{X}^2}{\Sigma X_i^2 - n\overline{X}^2} \right] \right) = S^2(a)$$

$$b \rightarrow N\left(B ; \frac{S^2_{yx}}{\Sigma X^2_i - n\overline{X}^2} \right) = S^2(b)$$



$$S = \sum (y_i - \hat{y}_i)^2$$

$\longrightarrow \hat{y}_i = a + bx_i$

si hacemos mínimos los errores $\longrightarrow a, b$

S^2_{yx} = " estimador de la varianza de los residuos " $= \dfrac{\sum (Y_i - \hat{Y}_i)^2}{n-2}$

S_{yx} = error estandar de estimación. Estima la dispersión de las observaciones alrededor de la recta de regresión muestral. Es una medida de la bondad del ajuste de la recta a la nube de puntos .

* Intervalos de confianza de los parametros del modelo . A y B .

B no puede ser conocido mientras no se observe toda la población su estimador puntual será b.

I.C. del (1 - α) % : $b \pm t_{\alpha/2;\,n-2} \cdot S(b)$

analogamente para A

I.C. del (1 - α) % : $a \pm t_{\alpha/2;\,n-2} \cdot S(a)$

* Prueba de hipótesis sobre la pendiente de la recta . (B) . Prueba de linealidad .

1) $H_O: B = 0$; $H_A: B \neq 0$

2) $H_O: B \leq 0$; $H_A: B > 0$

3) $H_O: B \geq 0$; $H_A: B < 0$

Se formulará la 2) cuando b salga + ; 3) si b sale -
pero ver si es suficientemente mayor que 0

El estádistico de contraste para la prueba sera $E.C. = \dfrac{b - B}{S(b)} = \dfrac{b}{S(b)}$

Si se considera B = 0 puede significar que :

- no existe relación lineal entre ambas variables . La var. considerada no es explicativa.
- existe relación pero esta no es lineal .(exponencial, logarítmica)

* Coeficiente de determinación r^2

Variabilidad total = $V.T. = \sum (Y_i - \overline{Y})^2$

Variabilidad explicada = $V.E. = \sum (\hat{Y}_i - \overline{Y})^2$

Variabilidad no explicada = $V.NE. = \sum (Y_i - \hat{Y}_i)^2$

V.T. = V.E. + V.NE.

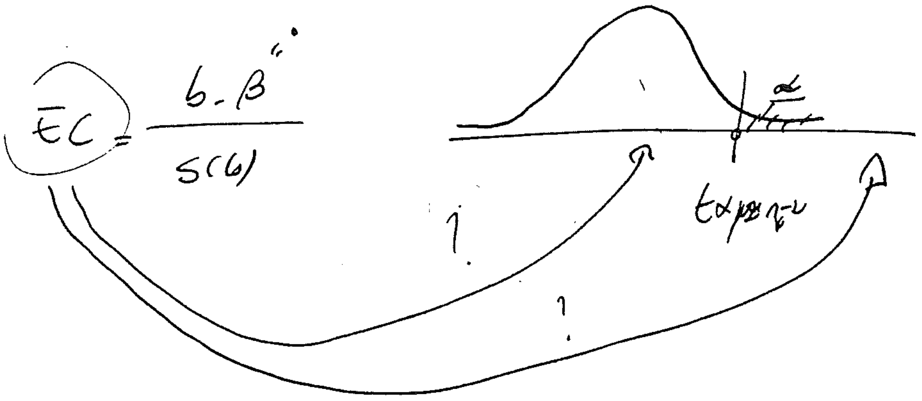
$r^2 = \dfrac{V.E.}{V.T.} = \dfrac{\sum (\hat{Y}_i - \overline{Y})^2}{S_y^2 \cdot n-1}$

r^2 mide el % de la var. total explicada por la recta de regresión .

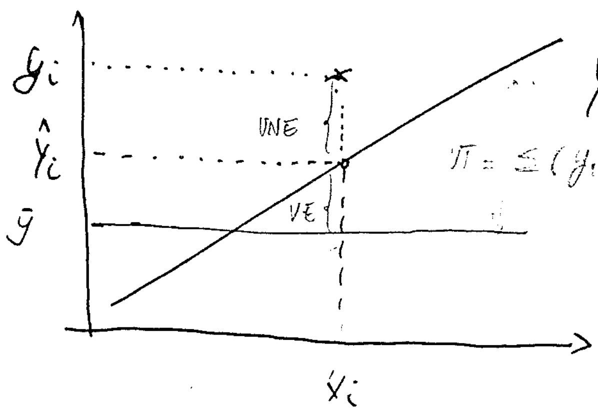
Mide el grado de asociación lineal de modo que si :

$r^2 = 1$ se tendrá una relación lineal perfecta .

finalmente
si $B>0 \rightarrow$ relación directa : $\Delta B \rightarrow \Delta Y_i$
si $B<0 \rightarrow$ " inversa : $\Delta B \rightarrow \nabla Y_i$



r^2 : coeficiente de determinación



Calculadora

$VT = \Sigma(y_i - \bar{y})^2 = S_y^2(n-1)$; $S_y^2 = \frac{\Sigma(y_i - \bar{y})^2}{n-1}$

$VE = \Sigma(\hat{y}_i - \bar{y})^2$
$VNE = \Sigma(y_i - \hat{y}_i)^2$ } $VT = VE + VNE$

cuanto mas sea mayor, mayor será la relación entre ambas variables.

VT tiene 2 componentes : VNE + VE

$r^2 = \frac{VE}{VT}$: idea que respecto al total (%)
cuanto más se acerque a 1 mejor.

Calculadora

Conociendo r^2 (de calc).

$r^2 = \frac{\Sigma(\hat{y}_i - \bar{y})^2}{S_y^2(n-1)} \rightarrow \Sigma(\hat{y}_i - \bar{y})^2 = r^2 \cdot S_y^2(n-1) = V.E$ } $VNE = VT - VE =$

$VT = S_y^2(n-1)$

conociendo (calc)

$= S_y^2(n-1) - (r^2 S_y^2(n-1)) = S_y^2(n-1)(1-r^2) = VN$

$\rightarrow S^2_{y/x} = \frac{VNE}{n-2} = \frac{S_y^2(n-1)(1-r^2)}{(n-2)}$ (q es + importante para la prueba de linealidad)

$r^2 = 0$ ausencia de relación lineal .

* Coeficiente de correlación . (r)

$r = \pm \sqrt{r^2}$ + si $b > 0$ y - si $b < 0$.

Tiene el mismo significado que r^2 . Si $r = 0$ no indica independencia sino únicamente ausencia de asociación lineal .

Previsiones .

$\hat{Y}_{n+1}$ = estimación del valor de Y para el valor de X_{n+1}

Y_{n+1} = verdadero valor de Y cuando X toma el valor X_{n+1}

Estimación puntual : $Y_{n+1} = \hat{Y}_{n+1}$

Estimación por intervalo de confianza : $\hat{Y}_{n+1} \pm t_{\alpha/2}\, S_p$

S_p es el estimador de la varianza del error de previsión .

$$S_p = S_{yx} \cdot \sqrt{1 + 1/n + \frac{(X_{n+1} - \overline{X})^2}{\sum X_i^2 - n\overline{X}^2}}$$

Se almacena materia prima utilizada en la producción de una fibra sintetica en un local sin control de humedad . Las medidas de humedad relativa en el lugar de almacenamiento y el contenido de humedad de una muestra del material (ambos en %) en ocho dias elegidos al azar mostraron los resultados siguientes :

Humedad	x	46	30	34	52	38	44	40	45
Contenido de humedad	y	10	7	9	13	8	12	11	11

a) Estimar los parámetros de la recta de ajuste .

b) Contrastar la hipótesis de linealidad .

c) Calcular e interpretar el coef. de determinación .

d) ¿ Cual será el contenido de humedad de la materia prima almacenada cuando la humedad en el local de almacenaje es del 50 % .

S_y^2 = 00'125

r^2 : 0'762

$\Sigma(x_i^2) - n\bar{x}^2$:

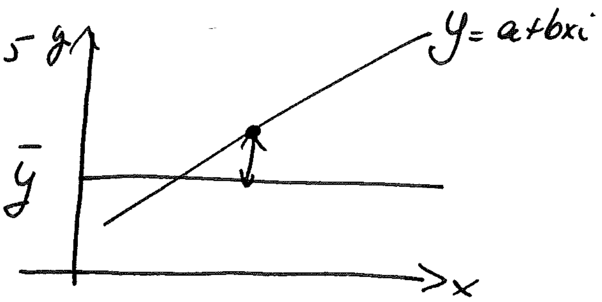
1.-

$$Y_i = A + Bx_i + E_i$$

$$\hat{y}_i = a + bx_i = -0'1745 - 0'25\,x_i$$

$$a = \frac{\sum_{i=1}^{n} y_i}{n} - b\,\frac{\sum x_i}{n} = -0'1745$$

$$b = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2} = 0'25$$



$$r^2 = 0'7621 \Rightarrow \frac{VE}{VT}$$

$H_0: B_1 \le 0$

$H_a: B_1 > 0$

$$E.C = \frac{b - B_0}{sb} = \frac{0'25}{sb} = 76'68$$

sb

$r^2 = 0'7621 \Rightarrow$

$$\begin{cases} r^2 = \frac{VE}{VT} \\ s_y^2 = \frac{\sum (y_i - \bar{y})^2}{(n-1)} = VT \end{cases} \rightarrow VT = s_y^2 \cdot (n-1) = 2'03^2 \cdot 7 = 28'875$$

$$r^2 = \frac{VE}{VT} = 0'7621 \rightarrow VE = 22$$

$X_{MH} = 50$

$$VNE = 28'875 - 22 = 6'875$$

$$VE \longrightarrow VNE = \sum (y_i - \hat{y}_i)^2$$

6±

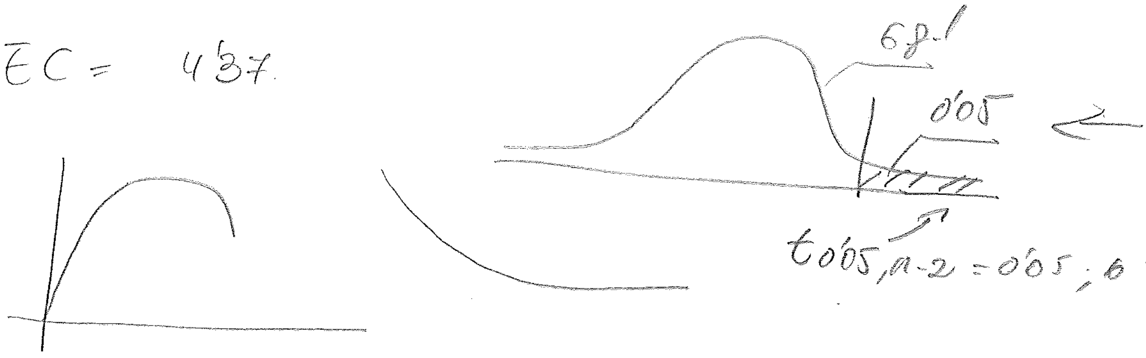
$$S_{yx}^2 = \frac{6'875}{6} = 1'1458$$

$$S_b^2 = \frac{1'1458}{13881 - 8 \cdot 41'125^2} = 0'00326555$$

13530'125

350'875

EC = 4'37



$t_{0'05, n-2} = 0'05; 6 = 1'94$

Rechazamos H_0

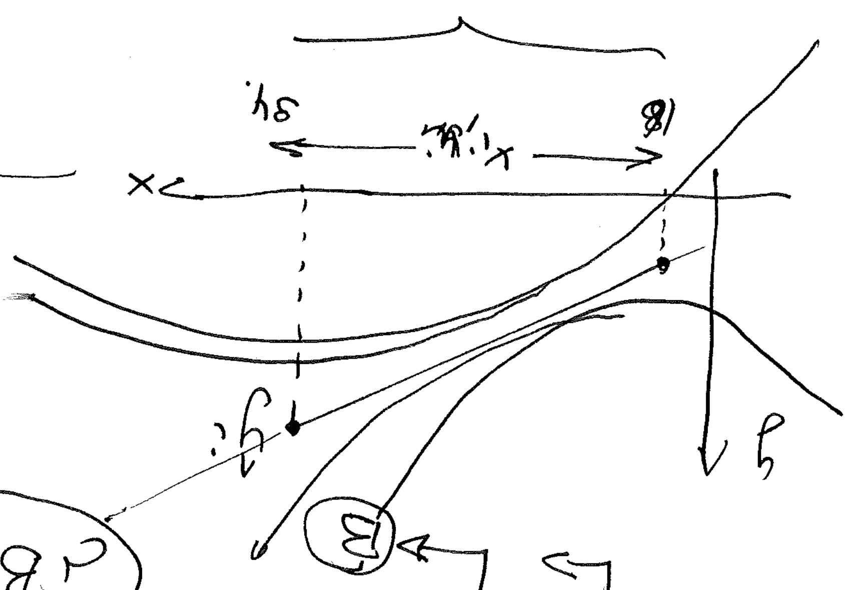
VT =

$S_y^2 = \frac{VT}{n-1}$

$VT = S_y^2 (n-1)$

~~A = P~~

$y = A + Bx_i \rightarrow \phi$



$y_i = A + Bx_i + E.r.$

$\Delta B = 2'5$

$y = A + Bx_i$

$A \times$ (nuevo) $\rightarrow \Delta y = 2'5$

```
-> REGRESSION
->    /STATISTICS COEFF OUTS CI R ANOVA
->    /CRITERIA=PIN(.05) POUT(.10)
->    /DEPENDENT conthume
->    /METHOD=ENTER hume  .
```

*** * * *   M U L T I P L E   R E G R E S S I O N   * * * ***

Listwise Deletion of Missing Data

Equation Number 1   Dependent Variable..   CONTHUME   contenido de humedad del *la materia prima*

Block Number 1. Method: Enter   HUME *humedad del local*

Variable(s) Entered on Step Number
   1..   HUME   humedad de la mat. prima

Multiple R		,87303
R Square	*(r^2)*	,76218
Adjusted R Square		,72254
Standard Error		1,06982 *= S_{yx} = 1,07*

---------------------- Variables in the Equation ----------------------

Variable		B	SE B *(S_b)*	95% Confdnce Intrvl B		Beta
HUME	*b//*	,250445	,057113	,110695	,390195	,873028
(Constant)	*a//*	-,174564	2,379035	-5,995821	5,646694	

----------- in ------------

Variable		T	Sig T
HUME	*t c.*	4,385	,0046
(Constant)		-,073	,9439

→ cae a la dcha ⇒ está a. l.c.

x explica cuanto varía y_i x cada variación unitaria de x_i

Una empresa quiere comprobar que en su división de Sestao los gastos de mantenimiento no tienen nada que ver con el volumen de producción, para lo cual eligen aleatoriamente cinco semanas el último año ,obteniendose los resultados :

Y GTOS MANTEN.	80	83	78	81	78
X UNI.PRODUC.	340	355	320	384	341

Basandote en esta muestra y para un nivel de sign. del 5% , ¿ Que le dirias a la empresa respecto a lo que quiere comprobar ?

" RESPUESTA DE LOTUS "

SALIDA DE REGRESION :

CONSTANTE	60,287
ERR STD DE Y EST	1,897
R al cuadrado	0.3996
N° de Observaciones	5
Grados de libertad	3

Coeficiente (s) x	0,0566
Err Std de Coef.	0,0400

gtos mantenimiento = f (producción) ?

V. Explicada — V. explicativa

$a = 60'287$

$b = 0'0566$

$\hat{y}_i = 60'28 + 0'056 x_i$

$r^2 = 0'4 = \frac{V.\ Explicada}{V\ Total} = \frac{VE}{S_y^2(n-1)} \Rightarrow VE = 0'4 \cdot 4'5 \cdot 4 = 7'2$

VT

$VNE = VT - VE = 18 - 7'2 = 10'8$

↳ es posible q exista otra variable más relevante q esta q explique más la y.

$S_{yx}^2 = \frac{10'8}{n-2} = 3'6$

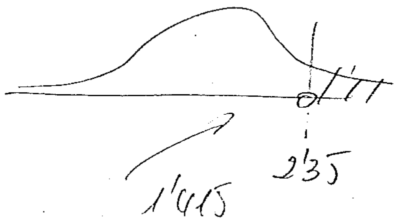
$S_b^2 = \frac{S_{yx}}{607162 - 5\cdot\bar{x}^2} = 1'6 \cdot 10^{-3}$

2242

$E_c = \frac{0'0566}{\sqrt{1'6\cdot 10^{-3}}} = \frac{0'0566}{0'04} = 1'415$

deja un area de ... a su ...

comparamos con $t_{\alpha; n-1}$ q nos deja a su dcha



Aceptamos H_0.

Prod no es una V. explicativa de los ~~Aumedi~~ gtos mant.

EXAMEN ENERO 1996

EN LA TABLA ADJUNTA SE PRESENTAN LOS DATOS MUESTRALES SOBRE EL NUMERO DE HORAS DE ESTUDIO INVERTIDAS POR OCHO ESTUDIANTES FUERA DE CLASE , DURANTE UN PERIODO DE TRES SEMANAS PARA EL CURSO DE ESTADISTICA II JUNTO CON LAS CALIFICACIONES QUE OBTUVIERON EN UN EXAMEN AL FINAL DE ESTE PERIODO .

ESTUDIANTE	1	2	3	4	5	6	7	8	
HORAS ESTUDIO	20	16	34	23	27	32	18	22	X
CALIFICACION	64	61	84	70	88	92	72	77	Y

a) ¿ QUE SE PODRÍA DECIR RESPECTO A LA RELACIÓN ENTRE HORAS DE ESTUDIO Y LA CALIFICACIÓN OBTENIDA ? . ¿ EXISTE MUCHA RELACIÓN ENTRE AMBAS VARIABLES ?

b) ¿ QUE CALIFICACIÓN OBTENDRÍA UN ESTUDIANTE QUE HUBIERA DEDICADO 15 HORAS A ESTUDIAR FUERA DE CLASE ?.

c) EL PROFESOR DE ESTÁDISTICA ESTÁ CONVENCIDO DE QUE AUMENTANDO EN UNA HORA LAS HORAS DE ESTUDIO LA CALIFICACIÓN MEJORA EN 2,5 . ¿ TIENE RAZÓN EN SU AFIRMACIÓN ? (NIVEL DE SIGN. PARA TODO EL PROBLEMA 1%)

METODOS DE PREVISIÓN

MODELO LINEAL DE REGRESIÓN SIMPLE.

humedad local = x $\quad \bar{x} = 41'125$

calculado de humedad : y

[MODE LR]

cont hum y p = f(hum local)
v. expl Y $\quad$ V explicativa X

$S_Y^2 = 4'125$

$a = -0'17456$

$b = 0'2504$

$r^2 = 0'762 = \frac{V Ex}{VT} = \frac{V.E}{VT} = \frac{VE}{S_y^2(n-1)} \Rightarrow VE = r^2 \cdot S_y^2 (n-1) =$

$= 0'762 \cdot 4'125 \cdot 7 = 22 = \Sigma()$

$S_y^2 = \frac{\Sigma(y_i - \bar{y})^2}{n-1}$ VT $\Rightarrow VT = S_y^2(n-1)$

$VT = VE + VNE \rightarrow VNE = 28'875 - 22 = 6'875 = VNE$

$VNE = \Sigma(y_i - \hat{y})^2$

$S_{yx}^2 = \frac{6'875}{n-2} = \frac{6'875}{6} = 1'1458$

$S(b)^2 = \frac{S_{yx}^2}{\Sigma x_i^2 - n\bar{x}^2} = \frac{1'146}{13881 - 8 \cdot 41'125^2} = \frac{1'146}{350'875} = 0'00326.$

RCL 1

$\hat{y}_i = -0'174 + 0'250 x_i$

$IC : (b) = 0'250 \pm t_{0'025;\, n-2} \cdot S(b) = 0'250 \pm 0'14$

$0'44 \quad \sqrt{0'00326} = 0'0571$

este int. contendrá a B grande con una prob. del 95%.

→ regla $\hat{y}_i$ uso de $B>0$

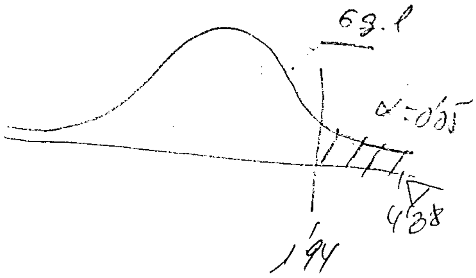
b) $H_0\ B \leq 0$

$H_a\ B > 0$

$$EC = \frac{b - B}{S(b)} = \frac{0'25}{0'057} = 4'38$$

vemos la necesidad de calcular antes

$VNE \longrightarrow S_{yx} \longrightarrow S(b)$



Rechazamos Ho; variable x es una variable relevante, es una variable explicativa del modelo.

c) $$r^2 = 0'762 = \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2} = \frac{VEx}{VT}$$: el 76'2% de las variaciones observadas de la vble y están explicadas por las variaciones de la vble X.

El 23'8% de las variaciones restantes se deben a la influencia de otras variables no incluidas en el modelo.

d) $x_{n+1} = 50$

$\hat{y}_{n+1} = -0'174 + 0'250 \cdot 50 = 12'326$ (estimación puntual)

mediante IC $\quad -0'174$: centro

$$t_{\alpha/2} \cdot S_{yx} \sqrt{1 + \frac{1}{n} + \frac{(x_{n+1} - \bar{x})^2}{\sum x_i^2 - n\bar{x}^2}} = 2'44 \cdot 1'07 \cdot \sqrt{1 + \frac{1}{8} + \frac{50}{350}}$$

$= 3'033 =$ Radio

I.C $\quad -0'174 \pm 3'033$

EXAMEN ENERO 96

$Y_i = \alpha + BX_i + \epsilon \rightarrow \hat{y}_i = a + bx_i = [illegible] + 1'4965 x_i$

$a = [illegible]$

$b = 1'4965$ [illegible]

$r^2 = 0'742 = \frac{VEX}{VT} = \frac{VE}{S_y^2(n-1)} \rightarrow VE = r^2 S_y^2 (n-1) = 0'743 \cdot 126'57 \cdot 7 = 658'5 = \Sigma(\hat{y}_i - \bar{y})^2 = VE$

$VT = 886$

$VNE = VT - VE = 886 - 658'5 = 227'5 = VNE$

$S^2_{y \cdot x} = \frac{VNE}{n-2} = \frac{227'5}{6} = 37'92$

$\hat{\sigma}^2(b) = \frac{S^2_{y \cdot x}}{\Sigma x_i^2 - n\bar{x}^2} = \frac{37'92}{4662 - 4068} = \frac{37'92}{594} = [illegible]$

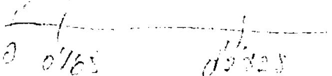
[illegible] de las variaciones observadas en la calificación vienen motivadas por las horas de estudio. El [illegible]% restante está motivado por influencia de otras variables no incluidas en el problema.

IC (β) $b \pm t_{\frac{\alpha}{2}; n-2} \cdot \hat{\sigma}_b$ · ~~precisión~~

$1'4966 \pm t_{0'005;6} \cdot \sqrt{[illegible]} = 1'4966 \pm 1'3315 \rightarrow P[0'165 < \beta < 2'828] = 0'99$

En el 99% de las poblac. B estará en este intervalo.

→ el IC no obtiene un valor nulo de β



para esa B no será nunca β

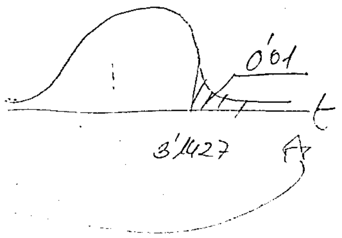
como $b > 0$

prueba de linealidad

$H_0 \; B \leq 0$

$H_a \; B > 0$

$E_C = \frac{b}{\hat{\sigma}(b)} = \frac{1'4965}{0'359} = 4'166$



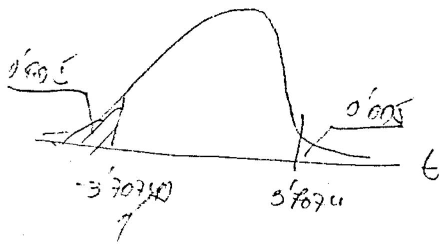
Rechazamos H_0

Aceptamos H_a ⟹ es explicativo.

c) $H_0 \; B = 2'5$

$H_a \; B \neq 2'5$

$E_C = \frac{1'4965 - 2'5}{0'359} = -2'794$



$t_{0'005,6} = 3'7074$

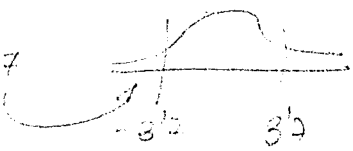
$t_{0'995,6} = -3'7074$

como el IC ya sabía q 2'5 está dentro luego aceptamos H_0

Aceptamos H_0. $B = 2'5$ en el 99% de los casos.

b) $\hat{y}_{15} = [illegible] + 1'4966 \cdot 15 = [illegible]$

podemos decir q $\mu_B = 3'7$

$EC = -4'187$  <u>es menor</u>.

REGRESIÓN REPASO

Se ha pasado un test de eficiencia lectora (TEL) a un conjunto de 22 alumnos universitarios escogidos aleatoriamente midiendoles la velocidad de lectura (palab./min) obteniendose los siguientes resultados :

X H	240	168	290	210	150	180	200	396	224	324	278
Y M	182	250	240	150	402	300	225	324	287	245	245

Para un nivel de sign. del 10% realizar los siguientes contrastes :

a) La velocidad media de lectura en los hombres es igual o mayor de 250 pal / min . (palabras)

b) La varianza del colectivo de hombres es de 4.900 $pal^2 / min.^2$.

c) La velocidad de lectura es independiente del sexo .

d) La velocidad media de lectura es la misma para hombres y mujeres .

¿ A partir de que diferencia muestral se considerará que las velocidades medias son distintas ? .

e) ¿ Se puede decir que los datos muestrales del colectivo de hombres provienen de una distribución normal ? .

→ Kolmogorov Smirnov.

REPASO

$\bar{x}_H = 241'81 \rightarrow S_H = 74'12 \rightarrow S^2 = 5493'96 \qquad n_x = 11$

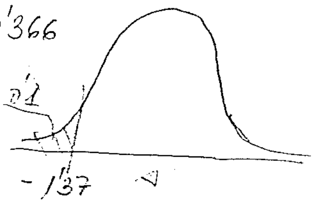
$\bar{x}_H = 259'1 \rightarrow S_H = 68'512 \rightarrow S^2 = 4693'89 \qquad n_y = 11$

$H_0: \mu_x \geq 250$

$H_a: \mu_x < 250$

$$EC = \frac{\bar{x} - \mu_x}{S/\sqrt{n}} \nearrow t_{n-1} = \frac{241'81 - 250}{74'12/\sqrt{11}} = -0'366$$

Valores crítico $t_{0'1;\,10}$: $1'8125$



Aceptamos Ho, la velocidad media en un 90% de los [illegible] será mayor o igual a 250 pal/min. Cometemos un error de tipo II ... β de aceptar Ho cuando esta sea falsa.

a) v.a x: Velocidad de lectura (pal/minuto) de un alumno
v.a y: " " " " " " una alumna.

Se supone, Nat (x) → Normal
Nat (y) → Normal

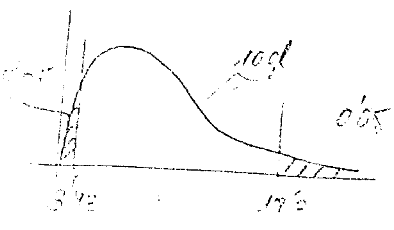
$\bar{X}$: velocidad media de lectura de un alumno de la muestra
$\bar{Y}$: " " " " " una alumna " " "

NOTA: si fuera bilateral la $t \rightarrow t_{(n-1);\,\alpha/2}$

b) $H_0: \sigma_x^2 = 4900$

$H_a: \sigma_x^2 \neq 4900$

$$EC \; \frac{S_x^2(n-1)}{\sigma_x^2} \longrightarrow \chi^2_{n-1} \;\Big\|\; EC = \frac{10 \cdot 74'12}{4900} = 15'21$$



Aceptamos Ho la varianza es 4900.

c) Tablas de contingencia o bondad de ajuste no, xq tenemos dos características cuantitativas naturales.
Diferencia de medias tampoco
Regresión tampoco.

entonces ⟶ pares apareados ⟶ $\bar{d}$: si las diferencias son pequeñas es q es independiente del sexo.

$\bar{d} = -76$

$S_d^2 = 104/97 = 11020'4$

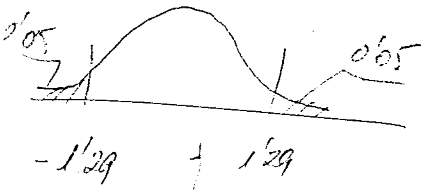
$\bar{d} = -17'27$ → vemos q el de las mujeres es mayor

$S^2_d = 10791'81$

$d_1 = 58$
$d_2 = -82$
$d_3 = 50$
$d_4 = 60$
$d_5 = -252$
$d_6 = -120$
$d_7 = -25$
$d_8 = 72$
$d_9 = -67$
$d_{10} = 79$
$d_{11} = 83$

$H_0 \; \mu_d = 0$

$H_a \; \mu_d \neq 0$



$$EC = \frac{\bar{d} - \mu_d}{S_d/\sqrt{n}} = \frac{-17'27 - 0}{103'88/\sqrt{11}} = -0'55$$

Aceptamos H_0 lo que significa que la velocidad de lectura es independiente del sexo. Podríamos realizar o cometer un error de Aceptar H_0 si esta es falsa.

d)

$H_0 : \mu_x - \mu_y = 0$

$H_a : \mu_x - \mu_y \neq 0$ $\qquad EC = ?$

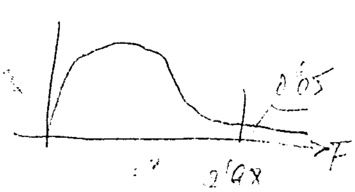
Prueba de homocedasticidad

$H_0 : \sigma_x^2 = \sigma_y^2$

$H_a : \sigma_x^2 \neq \sigma_y^2$

$EC = \frac{S_x^2}{S_y^2} = 1'17 \longrightarrow F_{10;10;0'05}$

son iguales



P(Aceptar H_0 / H_0 falsa) => $\alpha/2$ = $P(\bar{x} - \bar{y} > 2'4 \,/\, \mu_x - \mu_y = 5'7)$

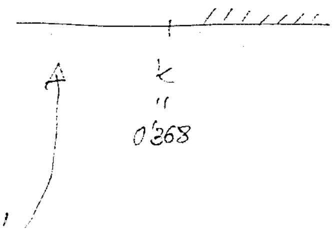
e)

		$S_n(x)$	F_{0x}	$F_{0x} - S_{nx}$	$F_{0x} - S_{n-1}x$
1	150	1/11 = 0'0909	0'1075	0'0166	0'1075
2	168	2/11 = 0'1818	0'1611	0'0207	0'1611 - 0'0909
3	180	3/11 = 0'27	0'233	0'03972	
4	200	4/11 = 0'36	0'2877	0'076	
5	210	5/11 = 0'45	0'3336	0'121	
6	224	6/11 = [illegible]	0'4052	0'14005	
7	240	7/11 = 0'63	0'4952	0'14106	
8	278	8/11 = 0'72	0'6737	0'03733	
9	290	9/11 = 0'81	0'7422	-0'076	
10	324	10/11 = 0'90	0'8665	0'0426	
11	296	11/11 = 1	0'9817	0'0198	

$\bar{x}$ = 241'81

S = 47'12

P(x < 150) = P(Z < -1'238) = 0'1075
P(x < 168) = P(Z < -0'996) = 0'1611
P(x < 180) = P(Z < -0'73) = 0'233
P(x < 200) = P(Z < -0'56) = 0'2877
P(x < 210) = P(Z < -0'43) = 0'3336
P(x < 224) = P(Z < -0'24) = 0'4052
P(x < 240) = P(Z < 0'024) = 0'4912
P(x < 278) = P(Z < 0'473) = 0'6737
P(x < 290) = P(Z < 0'65) = 0'7422
P(x < 324) = P(Z < 1'108) = 0'8665
P(x < 296) = P(Z < 2'03) = 0'9812



CC = 0'14106

Valor crítico : $k_{S_{0'9},10}$ = 0'368

Formulario: σ_x, σ_y → independientes y desconocidas

$$tc = \frac{(\bar{x}-\bar{y})-(\mu_x-\mu_y)}{S_p\sqrt{1/n_x+1/n_y}} \longrightarrow t_{n_x+n_y-2}$$

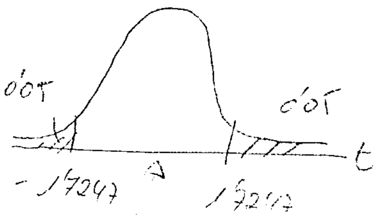
$$S_p^2 = \frac{(n_x-1)S_x^2+(n_y-1)S_y^2}{n_x+n_y-2} = \frac{10\cdot 74'12^2+10\cdot 68'81^2}{20} = 5093'697$$

$S_p = 71'37$.

$$tc = \frac{241'81-259'1}{71'37\sqrt{1/11+1/11}} = \frac{-17'29}{30'43} = -0'568$$

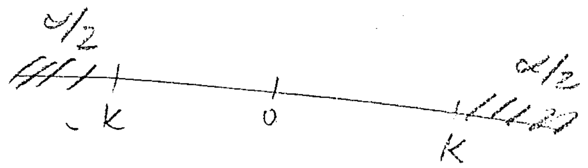
$t_{n_x+n_y-2} = t_{20;\,0'05} = 1'7247.$

$-t_{20;\,0'95} = -1'7247.$



Aceptamos Ho, las medias de ambos colectivos son las mismas

Nos piden el valor de K



$\alpha/2 = P(\text{rechazar Ho/cierta}) = P(\bar{x}-\bar{y} > k \,/\, \text{Ho cierta} = \mu_x-\mu_y=0) =$

$= P\left(t > \frac{k-0}{71'37\sqrt{2/11}}\right) = \alpha/2$

↓

$t_{\alpha/2;\,20} = 1'7247$ $\quad \frac{k}{71'37\sqrt{2/11}} = 1'7247 \longrightarrow k = 52'4$

$+k = 52'4$

$-k = -52'4$

a partir de 52'4 se considera que las diferencias son significativas

05.09.97.

«MÉTODOS DE REGRESIÓN LINEAL»

PROBLEMA 1:

a) Estimo los parámetros de la recta de ajuste.

$Y_i = A + BX_i + E_i$

Como tengo datos de una muestra calcularé los parámetros muestrales:

$\hat{Y}_i = a + bx_i$

* $b = \dfrac{\sum_{i=1}^{n} X_iY_i - n\bar{X}\bar{Y}}{\sum_{i=1}^{n} X_i^2 - n\bar{X}^2}$; $b = 0'2504$

$a = \dfrac{\sum_{i=1}^{n} Y_i}{n} - b\dfrac{\sum_{i=1}^{n} X_i}{n}$; $a = -0'1745$

$r^2 = \dfrac{\sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^{n} (Y_i - \bar{Y})^2}$; $r^2 = 0'762$

$r^2 = \dfrac{V.E}{V.T}$, $V.E = r^2 VT$

* $S_y^2 = 4'125$

$S_y^2 = \dfrac{\sum_{i=1}^{n} (Y_i - \bar{Y}_i)^2}{n-1}$, $V.T = S_y^2 \cdot (n-1)$

$V.T = 4'125 \cdot 7$; $V.T = 28'875$

$V.E = 0'762 \cdot 28'875$; $V.E = 22$ => $V.N.E = 6'875$

$S_{xy} = \sqrt{\dfrac{\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2}{n-2}}$; $S_{xy} = \sqrt{\dfrac{6'875}{6}}$, $S_{xy} = 1'0704$

$S_b = \dfrac{S_{xy}}{\sqrt{\sum X^2 - n\bar{X}^2}}$; $S_b = \dfrac{1'0704}{350'875}$; $S_b = 0'00326$

$\sum X^2 = 1388$

$n\bar{X}^2 = 1360'125$

$\hat{y}_i = -0'174 + 0'25x_i$

I.C. : (b) $= b \pm t_{\alpha/2,\, n-2} \cdot S_b = 0'25 \pm t_{0'025;6} \cdot 0'0571 = 0'25 \pm 0'14$

$P[\,0'11 \leq B \leq 0'29\,] = 95\%$

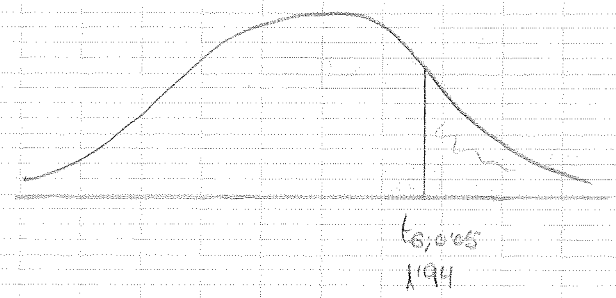
b) Contrasto la hipótesis de linealidad:

$H_0: B \leq 0$

$H_1: B > 0$

E.C.: $E.C. = \dfrac{b - B_0}{S_b} = \dfrac{0'025 - 0}{0'0571}$, $E.C. = 4'38$

V.C.: (unilateral por la dcha.)



Como E.C > V.C => Rechazamos H_0.

Para $\alpha = 5\%$ la v.a. X es relevante para los valores de la v.a. Y.

c) Calculo e interpreto el coef. de determinación:

$r^2 = 0'762$

El 76'2% de las variaciones de la v.a. Y vienen explicadas por las variaciones de la v.a. X.

d) Calculo el contenido en humedad de la m.p. almacenada cuando la humedad en el local de almacenaje es del 50%.

$x_i = 50$ $\quad \hat{y}_i = -0'174 + 0'25 \cdot 50 \Rightarrow \hat{y}_i = 12'326$

PROBLEMA 2:

$$Y_i = A + BX_i + E_i$$
$$\downarrow \quad \hat{Y}_i = a + bx_i$$

Calculo "a" y "b":

$$b = \frac{\sum_{i=1}^{n} X_iY_i - n\bar{X}\bar{Y}}{\sum_{i=1}^{n} X_i - n\bar{X}^2} \Rightarrow b = 0'0566$$

$$a = \frac{\sum_{i=1}^{n} Y_i}{n} - b\frac{\sum_{i=1}^{n} X_i}{n} \Rightarrow a = 60'287$$

$$r^2 = \frac{\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2} = \frac{V.E}{V.T} \Rightarrow r^2 = 0'4 \ ; \ V.E = r^2 \cdot V.T \ ; \ V.E = 7'2$$

$$* \ S_y^2 = \frac{\sum_{i=1}^{n}(Y_i - \bar{Y}_i)}{(n-1)} \Rightarrow 4'5 \cdot 4 = V.T \ ; \ V.T = 18.$$

$$V.N.E = 18 - 7'2 \ ; \ V.N.E = 10'8$$

$$S_{xy} = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{n-2}} \ ; \ S_{xy} = \frac{10'8}{3} \Rightarrow S_{xy} = 1'8973$$

$$S_b = \frac{S_{xy}}{\sqrt{\sum X^2 - n\bar{X}^2}} \ ; \ S_b = \frac{1'8973}{47'3497} \ ; \ S_b = 0'04$$

$$\hat{Y}_i = 60'287 + 0'0566X_i$$

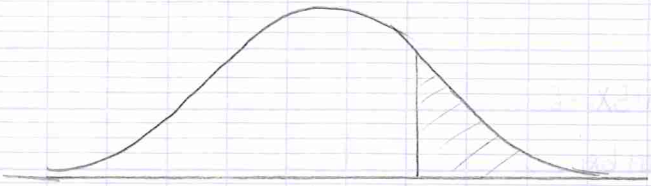
Hipótesis: H_0: $B \leq 0$.

H_1: $B > 0$.

E.C.:

$$E.C = \frac{b - B_0}{S_b} \ ; \ E.C = \frac{0'0566}{0'04} \ ; \ E.C. = 1'415$$

V.C.:



$t_{\alpha;n-2}$

2'3534

Como E.C. < V.C -> Aceptamos H_0.

Las unidades de producción no tiene nada que ver con los gtos. de mantenimiento.

PROBLEMA 3:

a) $Y_i = A + BX_i + E_i$

$\hat{Y}_i = a + bx_i$

$$b = \frac{\sum_{i=1}^{n} X_i Y_i - n\bar{X}\bar{Y}}{\sum_{i=1}^{n} X_i^2 - n\bar{X}^2} ; \quad \underline{b = 1'4965}$$

$$a = \frac{\sum_{i=1}^{n} Y_i}{n} - b\frac{\sum_{i=1}^{n} X_i}{n} ; \quad \underline{a = 40'08}$$

$$r^2 = \frac{\sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^{n} (Y_i - \bar{Y})^2} ; \quad \underline{r^2 = 0'7432}$$

$$S_y^2 = \frac{\sum_{i=1}^{n} (Y_i - \bar{Y})^2}{(n-1)} ; \quad V.T = S_y^2 \cdot (n-1) ; \quad \underline{V.T = 886}$$

($\sum_{i=1}^{n} (Y_i - \bar{Y})^2$ -> V.T)

$$r^2 = V.E / V.T ; \quad V.E = r^2 \cdot VT ; \quad \underline{V.E = 658'4752}$$

$$V.N.E = V.T - V.E ; \quad \underline{V.N.E = 227'5248}$$

$$S_{xy} = \sqrt{\frac{\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2}{n-2}} ; \quad S_{xy} = \sqrt{\frac{227'5248}{6}} ; \quad \underline{S_{yx} = 6'1579}$$

$$S_b = \frac{S_{xy}}{\sqrt{\sum X^2 - n\bar{X}^2}} ; \quad S_b = \frac{6'1579}{17'1464} ; \quad \underline{S_b = 0'36.}$$

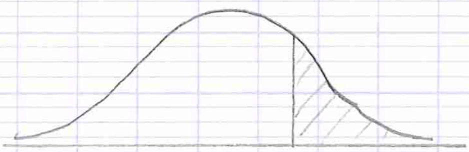
$\hat{Y}_i = 40'08 + 1'4965 X_i$

$H_0: B \leq 0.$

$H_1: B > 0.$

E.C.: $E.C. = \dfrac{b - B_0}{S_b}$; $E.C. = \dfrac{1'4965}{0'359}$; $E.C. = 4'1669$

V.C.:



$t_{0'01;6}$

3'1427

Como E.C. > V.C. rechazamos la H_0.

Para $\alpha = 1\%$ las horas de estudio justifican la calificación del estudiante.

Podemos afirmar que la relación que existe entre ambas vbles. tampoco es muy grande, ya que $r = 0'4 < 0'65$.

b) Calculo la calificación que obtendría un estudiante que dedicara 16 hrs:

$B = b \pm t_{n-2;\alpha/2} \cdot S_b$; $B = 1'4965 \pm 3'7074 \cdot 0'359$

$B = 1'4965 \pm 1'33$

c) Demostramos si el profesor tiene razón:

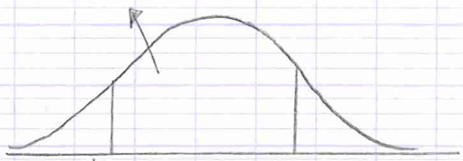
$\Delta X = 1$ hr.

$\Delta Y = 2'5 \Rightarrow B = 2'5$

$H_0: B = 2'5$

$H_1: B \neq 2'5$

E.C: $E.C = \dfrac{1'4965 - 2'5}{0'359}$; $E.C = -2'795$



$t_{0'005;6}$ $t_{0'005;6}$

-3'7074   3'7074

Aceptamos la H_0: El profesor tiene razón.

PROBLEMA 3:

Para $\alpha = 0'1$ realizo los siguientes contrastes:

a) velocidad media de lectura en hombres es ≥ 250 pal/min:

$\bar{X} = 241'81$; $S_x = 74'12$ $n_x = 11$

$\bar{Y} = 259'1$; $S_y = 68'512$ $n_y = 11$.

H_0: $\mu_x \geq 250$.

H_1: $\mu_x < 250$.

$n = 11$

$\alpha = 10\%$

E.C.:

$$E.C = \frac{\bar{X} - \mu_x}{S/\sqrt{n}} ; \; E.C. = \frac{241'81 - 250}{74'12/\sqrt{11}} ; \; E.C. = -0'366$$

V.C.: (prueba unilateral por la izq.)

$-t_{0'1;10} = -1'3722$

Si E.C. < V.C. Rechazamos H_0.

Como $-0'366 > -1'37'22$ Aceptamos H_0.

En el 90% de las ocasiones la velocidad media de lectura en hombres será mayor o igual a 250 pal/min.

b) la varianza del colectivo de hombres es de 4.900 pal²/min.²

H_0: $\sigma_x^2 = 4.900$

H_1: $\sigma_x^2 \neq 4.900$

E.C.:

$$E.C = \frac{(n-1)S_x^2}{\sigma_x^2} : \chi^2_{n-1} ; \; E.C = \frac{10 \cdot 5493'7714}{4.900} , \; E.C = 11'21$$

V.C.:

$\chi^2_{10;0'05} = 3'93$

$\chi^2_{10;0'95} = 18'3$

Como $3'93 \leq 11'21 \leq 18'3$ Acepto H_0.

En el 90% de las ocasiones la $\sigma_x^2 = 4.900$.

c) La velocidad de la lectura es independiente del sexo:

Calculamos la diferencia de velocidad, si la diferencia es pequeña la velocidad no depende del sexo.

$d_1 = 58$ $d_2 = -82$ $d_3 = 50$ $d_4 = 60$ $d_5 = -252$ $d_6 = -120$ $d_7 = -25$ $d_8 = 78$ $d_9 = -67$ $d_{10} = 79$

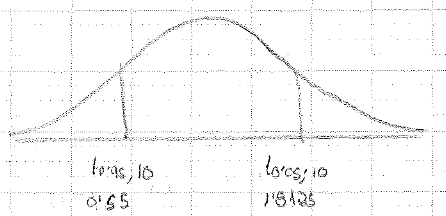
$d_{11} = 33$. $\bar{d} = -5'4545$; $S_d = 105'5427$

$H_0: \mu_d = 0$

$H_1: \mu_d \neq 0$

E.C.:

$$E.C. = \frac{-5'4545 - 0}{105'5427/\sqrt{11}} \; ; \; E.C. = -0'1714$$



Aceptamos la H_0: la velocidad de lectura es independiente del sexo.

d) La velocidad media de lectura es la misma para hombres y mujeres:

$H_0: \mu_x - \mu_y = 0$.

$H_1: \mu_x - \mu_y \neq 0$.

Prueba de homocedasticidad:

$H_0: \sigma_x^2 = \sigma_y^2$

$H_1: \sigma_x^2 \neq \sigma_y^2$

$E.C.: \frac{S_y^2}{S_x^2} = 1'17 \longrightarrow F_{10,10;0'05} = 2'98$ Como E.C < V.C Aceptamos H_0.

Para $\alpha = 10\%$ aceptamos que la velocidad media es igual para ambos sexos.

e) Observamos si los datos muestrales del colectivo de hombres provienen de una distribución normal. (Kolmogorov - Smirnoff).

	X	$S_{11}(x)$	$F_0(x)$	$F_0(x) - S_{n-1}(x)$
(150)	1	$1/11 = 0'0909$	0'1075	0'1075
(168)	2	$2/11 = 0'1818$	0'1611	0'0702
(180)	3	$3/11 = 0'\widehat{27}$	0'233	0'053
(200)	4	$4/11 = 0'\widehat{36}$	0'2877	0'0177
(210)	5	$5/11 = 0'\widehat{45}$	0'3386	-0'024
(224)	6	$6/11 = 0'\widehat{54}$	0'4052	-0'0448
(240)	7	$7/11 = 0'\widehat{63}$	0'4952	0'0448
(278)	8	$8/11 = 0'\widehat{72}$	0'6737	0'0437
(290)	9	$9/11 = 0'\widehat{81}$	0'7422	
(324)	10	$10/11 = 0'\widehat{90}$	0'8665	
(396)	11	$11/11 = 1$	0'9212	

$\bar{X} = 241'\widehat{81}$; $S_x = 74'12$